

Clinical Research

What Are the Minimal and Substantial Improvements in the HOOS and KOOS and JR Versions After Total Joint Replacement?

Stephen Lyman PhD, Yuo-Yu Lee MS, Alexander S. McLawhorn MD, MBA, Wasif Islam BS, Catherine H. MacLean MD, PhD

Received: 3 April 2018 / Accepted: 27 July 2018 / Published online: 1 September 2018
Copyright © 2018 by the Association of Bone and Joint Surgeons

Abstract

Background Patient-reported outcome measures (PROMs) are a gold standard for measuring therapeutic outcomes in research. Extending their use to inform clinical care decisions, determine the appropriateness of therapeutic choices, and assess healthcare quality is attractive but will require our professional community to establish valid estimates of minimal and substantial clinical improvements.

Questions/purposes The purposes of this study were (1) to assess the validity of estimates for the minimal clinically important difference (MCID) calculated using distribution-

and anchor-based methods by determining whether they exceed the minimal detectable change (MDC) for the Hip Disability and Osteoarthritis Outcome Score (HOOS) and the Knee Injury and Osteoarthritis Outcome Score (KOOS) domains, the HOOS, joint replacement (JR) and the KOOS, JR among patients who underwent THA or TKA; (2) to determine substantial clinical benefit thresholds for the HOOS and KOOS domains, the HOOS, JR, and the KOOS, JR among patients who underwent THA or TKA; and (3) to assess the proportions of patients who underwent THA or TKA who achieved an MCID for the HOOS and

One of the authors certifies that he (SL) received research support funding in the amount of more than USD 1,000,001 from the National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases. One of the authors certifies that he (SL) has received or may receive payments or benefits, during the study period, from the *Journal of Bone and Joint Surgery* (Needham, MA, USA) in the amount of USD 10,001 to USD 100,000; the Japanese Orthopedic Society of Knee Arthroscopy and Sports Medicine (Tokyo, Japan) in the amount of less than USD 10,000; Omni, Inc (Raynham, MA, USA) in the amount of less than USD 10,000; and Universal Research Solutions (Columbia, MO, USA) in the amount of USD 10,001 to USD 100,000, all outside this submitted work. One of the authors certifies that he (ASM), or a member of his immediate family, has received or may receive payments or benefits, an amount of less than USD 10,000 from Ethicon (Somerville, NJ, USA), an amount of less than USD 10,000 from Intellijoint (Waterloo, Ontario, Canada), and an amount of less than USD 10,000 from the Orthopaedic Research and Education Foundation (Rosemont, IL, USA), all outside this submitted work.

Clinical Orthopaedics and Related Research® neither advocates nor endorses the use of any treatment, drug, or device. Readers are encouraged to always seek additional information, including FDA approval status, of any drug or device before clinical use. Each author certifies that his or her institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

This work was performed at the Hospital for Special Surgery, New York, NY, USA.

S. Lyman, Y.-Y. Lee, Healthcare Research Institute, Hospital for Special Surgery, New York, NY, USA

A. S. McLawhorn, Department of Orthopedic Surgery, Hospital for Special Surgery, New York, NY, USA

W. Islam, C. H. MacLean, Center for the Advancement of Value in Musculoskeletal Care, Hospital for Special Surgery, New York, NY, USA

S. Lyman (✉), Hospital for Special Surgery, 535 E 70th Street, New York, NY 10021, USA, email: lymans@hss.edu

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

KOOS domains, HOOS, JR, and KOOS, JR based on distribution-based and anchor-based methods as well as the percentages of patients who achieved substantial clinical benefit using the anchor-based method.

Methods Medicare patients enrolled in our institutional joint replacement registry who subsequently underwent THA (n = 2323) or TKA (n = 2630) between 2007 and 2012 completed HOOS or KOOS preoperatively and 2 years postoperatively. Short-form joint replacement (JR) versions of each PROM were derived from the full PROMs. Of all eligible patients, 78% (3161 of 4080) of THAs and 74% of TKAs (3815 of 5156) consented to join the registry and completed a baseline survey, 88% (2796 of 3161) of THAs and 85% (3230 of 3815) of TKAs were eligible for followup survey administration, and 83% of THAs (2323 of 2796) and 81% (2630 of 3230) of TKAs returned 2-year surveys. For each HOOS domain, KOOS domain, HOOS, JR, and KOOS, JR, we calculated the calibration variation of the instrument (MDC) with confidence intervals (CIs) reflecting 80% (MDC80), 90% (MDC90), and 95% (MDC95) certainty; we calculated the smallest difference joint health patients might detect (MCID) using distribution- and anchor-based approaches and the difference that can be considered a large improvement in joint health (substantial clinical benefit) using an anchor-based approach.

Results Patients undergoing THA were 57% female with a mean (\pm SD) age of 73 ± 6 years, whereas patients undergoing TKA were 63% female with a mean age of 74 ± 6 years. Depending on the CI chosen for the MDC, values ranged from 7 to 16 for the HOOS and KOOS domains and the JRs. The MCIDs ranged from 6 to 9 for the distribution-based approach and 7 to 36 for the anchor-based approach. All HOOS and KOOS domains and all JR scores are scores from 0 (worst joint health) to 100 (best joint health). The MCIDs calculated using the distribution-based approach were not valid, because they were lower than the MDC for all HOOS/KOOS domains and both JRs at every confidence level. The anchor-based receiver operating characteristic approach, on the other hand, resulted in MCIDs exceeding MDC80 for seven of eight HOOS/KOOS domains and MDC95 for both JR scores. For all domains and JR versions, substantial clinical benefits ranged from 15 to 36, exceeding MDC95 in all domains and JR scores. Across HOOS and KOOS domains as well as the JR, the proportion of patients undergoing THA who achieved an MCID ranged from 77% to 95% with the distribution-based method and from 67% to 96% using the anchor-based method. The proportion achieving substantial clinical benefit ranged from 67% to 85%.

Conclusions The MDC and MCID differ greatly based on assumptions and methods used. The MCID anchor-based approach had superior construct and face validity compared with the MCID distribution-based approach, which

never exceeded even small MDCs. Achieving consensus about standard definitions of meaningful improvement will be necessary to maximize utility of these PROMs to inform clinical care or performance measurement.

Level of Evidence Level III, diagnostic study.

Introduction

Long held as an important, patient-centered approach for evaluating the effectiveness of elective orthopaedic procedures, patient-reported outcome measures (PROMs) are increasingly being recognized for their potential utility in care delivery and the assessment of care quality [21, 23, 25, 34]. The most well-validated joint-specific PROMs for arthroplasty are the Hip Disability and Osteoarthritis Outcomes Survey (HOOS) [28] and the Knee Injury and Osteoarthritis Outcomes Survey (KOOS) [32]. These PROMs along with their short-form versions for joint replacement (HOOS, JR and KOOS, JR, respectively) [16, 17] have been adopted by the Centers for Medicare & Medicaid Services (CMS) in their Comprehensive Care for Joint Replacement (CJR) Model, which is a bundled payment plan for primary elective total joint replacement. As part of this program, for the first time the CMS has incentivized PROM collection [8]. However, meaningful use of PROMs for assessing care delivery and/or care quality requires validated methodologies to measure changes in PROM scores and, based on that, an understanding of how much change over time should be expected and, most importantly, how much change matters to patients. These measures have not yet been established in our professional community.

Various methods can be used to calculate meaningful change in PROM scores such as the minimal detectable change (MDC), minimally clinically important difference (MCID), and substantial clinical benefit [1, 7, 29]. The MDC is the minimal amount of change in a PROM score required to signify a true health change given variability resulting from measurement error inherent in a given PROM [26]. The MCID, also sometimes referred to as the “minimal clinically important improvement,” reflects the minimum change in PROM scores that patients can perceive as a change in their health [2, 31, 37]. The substantial clinical benefit is the lower bound for defining optimal patient benefit [12]. Although the definitions are clear, methodology choices can result in wide variability in estimated values [30]. This variability poses a critical challenge, because these values carry important clinical implications; changes in patients’ scores that are lower than the MDC should be considered irrelevant. Changes in patients’ scores that are lower than the MCID imply treatment failures. Change in patients’ scores that fall between the MCID and substantial clinical benefit represent

changes that may range between the perceptible and the meaningful, but less than the substantial (depending on how the MCID is anchored in the surveys on which it is based). Only scores that exceed the substantial clinical benefit should be considered completely successful.

Using PROMs to inform care decisions, determine the appropriateness of therapeutic choices, and assess health care quality in orthopaedics requires an understanding of which calculations are most reliable and best reflect true improvement in joint-related health. Previous studies have attempted to establish MCID, MCD, and/or substantial clinical benefit for HOOS and KOOS, but their cohorts were not large, and the findings in these studies were not consistent with each other [3, 4, 22, 30, 35]. Validated measures for MCID, MCD, and substantial clinical benefit, which are agreed on by our professional community, are still lacking and needed if we are to be able to use HOOS, KOOS, or their JR versions accurately and fairly for clinical applications.

We therefore investigated our institution's extensive TKA and THA registries (1) to assess the validity of estimates for MCIDs calculated using distribution- and anchor-based methods by determining whether they exceeded the MDC for the HOOS and the KOOS domains, the HOOS, JR, and the KOOS, JR among patients who underwent THA or TKA; (2) to determine substantial clinical benefit thresholds for the HOOS and KOOS domains, the HOOS, JR, and the KOOS, JR among patients who underwent THA or TKA; and (3) to assess the proportions of patients who underwent THA or TKA who achieved an MCID for the HOOS and KOOS domains, HOOS, JR, and KOOS, JR based on distribution-based and anchor-based methods as well as the percentages of patients who achieved substantial clinical benefit using the anchor-based method.

Materials and Methods

This retrospective study draws data from a longitudinally maintained institutional registry; the Hospital for Special Surgery (HSS) Joint Replacement Registry includes all consenting patients undergoing THA and TKA between May 2007 and April 2012. Approximately 80% of consented patients completed HOOS or KOOS preoperatively as part of the requirement for inclusion in the registry and were mailed or emailed the same PROMs 2 years after their index surgery. The HOOS and KOOS are self-administered and completed without the aid of the surgeon or other medical staff.

All Medicare patients who met the CJR bundle inclusion criteria and who had completed preoperative and 2-year postoperative PROMs were included in this analysis [8]. Patients were ineligible if they underwent

revision or another THA or TKA within 2 years of their index surgery. In terms of data completeness, 78% (3161 of 4080) of THAs and 74% (3815 of 5156) of TKAs consented to join the registry and completed a baseline survey, 88% (2796 of 3161) of THAs and 85% (3230 of 3815) of TKAs were eligible for followup survey administration, and 83% (2323 of 2796) of THAs and 81% (2630 of 3230) of TKAs returned 2-year surveys. All patients underwent a primary, unilateral TKA or THA and reported pre- and postoperative PROMs.

We collected data from responses to the full HOOS and KOOS with all calculable osteoarthritis or total joint replacement-relevant domain scores, including pain, symptoms, activities of daily living (ADL), and quality of life (QOL). The JR versions were calculated from the full HOOS and KOOS. All HOOS and KOOS domains and all JR are scored from 0 (worst joint health) to 100 (best joint health).

In all, 2323 patients who underwent THA and 2630 patients who had TKA met our inclusion criteria. These Medicare patients with complete PROMs were comparable to all primary THA and TKA Medicare patients treated at our institution during the study period who also met criteria for inclusion in the CJR bundle (Table 1).

Statistical Analysis

We analyzed the HOOS responses from 2323 patients who had THA and the KOOS responses from 2630 patients who underwent TKA. All mean 2-year postoperative HOOS and KOOS scores were compared with the preoperative baseline scores (Table 2).

Minimal Detectable Change

The MDC is the minimal amount of change required to distinguish a true health change from variability resulting from measurement error; its calculations rely on a distribution-based approach that reflects a correction factor applied to the standard error of measurement (SEM) [26]. We calculated the MDC as follows: $MDC = z \text{ score} \times SEM \times \sqrt{2}$. We evaluated the MDCs z score with confidence intervals (CIs) reflecting 80%, 90%, and 95% certainty. The resulting MDC80, MDC90, and MDC95 correspond to a 5:1, 10:1, and 20:1 likelihood, respectively, of a patient's true PROM score falling outside the reported score. We applied a z score of 1.28 to MDC80, a z score of 1.64 to MDC90, and a z score of 1.96 to MDC95. We calculated SEM values as follows: $SEM = SD \times \sqrt{1 - ICC}$ with SD representing the SD of the baseline measurement and intraclass correlation coefficient (ICC) from the earlier studies [11, 33], which

Table 1. Demographics for THA and TKA cohorts and eligible patients at the study institution during the study period

Variables	THA study cohort (n = 2323)	All HSS THA Medicare patients (n = 7021)	TKA study cohort (n = 2630)	All HSS TKA Medicare patients (n = 8477)
Age (years)*	73 ± 6	74 ± 6	74 ± 6	74 ± 6
Women, number (%)	1395 (57%)	4249 (61%)	1740 (63%)	5656 (67%)
Body mass index* [†]	27 ± 5	27 ± 5	29 ± 6	30 ± 6
Obesity class, number (%)				
Underweight	31 (1%)	104 (2%)	7 (<1%)	38 (1%)
Normal weight	812 (33%)	2267 (34%)	633 (23%)	1736 (21%)
Overweight	983 (40%)	2612 (39%)	1035 (37%)	3006 (37%)
Obese–Class I	445 (18%)	1248 (19%)	663 (24%)	2062 (25%)
Obese–Class II	120 (5%)	384 (6%)	277 (10%)	886 (11%)
Obese–Class III	41 (2%)	147 (2%)	147 (5%)	440 (5%)
ASA class, number (%)				
I	38 (2%)	126 (2%)	37 (1%)	82 (1%)
II	1804 (74%)	5009 (71%)	1996 (72%)	5975 (71%)
III	588 (24%)	1865 (27%)	731 (26%)	2408 (28%)
IV	5 (< 1%)	16 (< 1%)	1 (< 1%)	9 (< 1%)
Charlson Comorbidity Index, number (%)				
0	1736 (71%)	4836 (69%)	1842 (67%)	5438 (64%)
1-2	600 (25%)	1913 (27%)	827 (30%)	2696 (32%)
3+	100 (4%)	272 (4%)	97 (4%)	343 (4%)

*Mean ± SD.

[†]body mass index (kg/m²); HSS = Hospital for Special Surgery; ASA = American Society of Anesthesiologists.

had a median reliability of 0.90 across all HOOS/KOOS domains.

Minimal Clinically Important Difference

The MCID, also sometimes referred to as the “minimal clinically important change” or “minimal clinically important improvement,” reflects the minimum change in PROM scores that a patient perceives as a change in their health [2, 9, 31, 37]. The MCID can be estimated using at least 14 methods [39]. For this analysis we compared distribution-based and anchor-based receiver operating characteristic (ROC) curve methods [4, 12, 30]. We did not assess consensus methods based on clinician opinions. We reasoned that estimates derived from such methods violate the intent of PROMs: to understand outcomes exclusively from the patient’s perspective [24].

The distribution-based approach assumes a normal response distribution. The calculation most often used is 0.5 x SD of the delta (change from baseline to followup) [4]. The anchor-based approach relies on an anchor question, asked at followup, and distinguishes patients who have had a change in their state of health from those who have not.

For our anchor questions, we selected the QOL satisfaction item from the HSS Satisfaction Survey at 2-year

followup (Fig. 1) because it is joint-specific and has previously been used as an anchor for MCID calculation in TKA [19].

Then, we used a logistic regression and the area under the ROC curve to find the delta score cut point that best identified which patients experienced a minimal improvement and those who did not according to anchor question responses.

To define minimal improvement for the anchor-based MCID, we considered the difference between those who answered “moderate improvement” (experienced minimal improvement) to those who reported “a little improvement” or “no improvement” (did not experience minimal improvement) on their QOL measure. We combined the little and no improvement groups because the no improvement group was small (Fig. 1).

Substantial Clinical Benefit

Substantial clinical benefit, which was initially proposed by Glassman et al. [12] to be the lower bound for defining optimal patient benefit, was calculated using the anchor-based ROC approach. Our anchor was again the QOL item from the HSS Satisfaction Survey. We compared those who answered “more improvement than I ever dreamed

Table 2. Baseline and 2-year PROM scores for THA and TKA cohorts

Domain	THA cohort					
	Baseline		2 years		Delta	
	Mean	SD	Mean	SD	Mean	SD
HOOS						
Pain	49	17	93	12	44	19
Symptoms	54	18	92	12	38	20
ADL	50	18	90	14	39	20
QOL	27	18	83	20	56	25
HOOS, JR	52	14	88	14	36	18
Domain	TKA cohort					
	Baseline		2 years		Delta	
	Mean	SD	Mean	SD	Mean	SD
KOOS						
Pain	51	17	88	15	37	20
Symptoms	55	18	83	15	28	22
ADL	55	17	85	16	30	19
QOL	28	17	72	24	44	26
KOOS, JR	53	13	80	14	28	17

PROM = patient-reported outcome measure; HOOS = Hip Disability and Osteoarthritis Outcome Score; ADL = activities of daily living; QOL = quality of life; JR = joint replacement short form of the HOOS and KOOS surveys; KOOS = Knee Injury and Osteoarthritis Outcome Score.

possible” and “great improvement” with those who reported “a little improvement” or “no improvement.” We combined the more and great improvement categories because the mean change for these groups was < 4 out of 100 points across all the evaluated PROM domains.

After calculating candidate MDC, MCID, and substantial clinical benefit values for all domains of the HOOS/KOOS and the HOOS/KOOS, JR, we determined how many patients in the cohort achieved these values. Finally, we considered the face validity of the various MDC, MCID, and substantial clinical benefit values based on conceptual frameworks requiring that MDC be less than MCID and that MCID be less than or equal to substantial clinical benefit. We reasoned that an MCID or substantial clinical benefit that was smaller than an MDC, which

represents the error of the scoring instrument, should not be considered a valid estimate [36].

Results

Validity of MCID Estimates

We assessed the validity of estimates for MCIDs calculated using distribution- and anchor-based methods by determining whether they exceeded the MDC for the HOOS, the KOOS, the HOOS, JR, and the KOOS, JR domains among patients who underwent THA or TKA. The MDC varied substantially based on selected CI, ranging from 7 to 16 for each HOOS and KOOS domain and the JR PROMs (Table 3). As

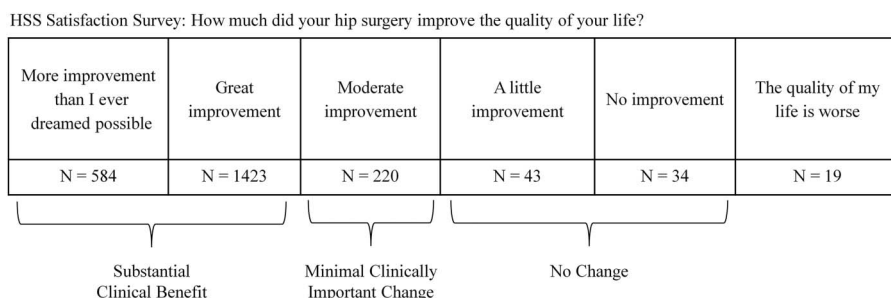


Fig. 1 The QOL item from the HSS Satisfaction Survey was used as the anchor question for calculating MCID and substantial clinical benefit, as described in the Materials and Methods.

Table 3. Minimal detectable change, minimal clinically important change, and substantial clinical benefit

Domains HOOS	Minimal detectable change			Minimal clinically important change		Substantial clinical benefit
	MDC80	MDC90	MDC95	Distribution-based	Anchor-based	Anchor-based
Pain	10	13	15	9	36	36
Symptoms	10	13	16	9	20	25
ADL	10	13	16	9	14	24
QOL	10	13	16	9	13	27
HOOS, JR	8	11	13	7	18	22

KOOS	Minimal detectable change			Minimal clinically important change		Substantial clinical benefit
	MDC80	MDC90	MDC95	Distribution-based	Anchor-based	Anchor-based
Pain	10	13	15	8	18	22
Symptoms	10	12	16	9	7	21
ADL	10	13	15	9	16	15
QOL	9	12	14	8	17	23
KOOS, JR	7	9	11	6	14	20

HOOS = Hip Disability and Osteoarthritis Outcome Score; MDC = minimal detectable change calculated with confidence intervals reflecting 80%, 90%, and 95% certainty reported, respectively, as MDC80, MDC90, and MDC95; ADL = activities of daily living; QOL = quality of life; JR = joint replacement short form of the HOOS and KOOS surveys; KOOS = Knee Injury and Osteoarthritis Outcome Score.

expected, values for the MDC95 were generally approximately 50% greater than those for the MDC80 across all domains for both HOOS and KOOS and the JR versions. The JR versions of the HOOS and the KOOS had smaller MDC values than any of the HOOS/KOOS domains.

Across the HOOS/KOOS domains and the JR versions, the MCIDs ranged from 6 to 9 for the distribution-based approach, and they ranged from 7 to 36 for the anchor-based approach.

The MCIDs calculated using the distribution-based approach were smaller than the corresponding MDC80, MDC90, and MDC95 for every domain and each JR version for both the HOOS and the KOOS, indicating that they were not valid (Table 3).

The MCIDs calculated using the anchor-based ROC were greater than the corresponding MDC80 and MDC90 for all HOOS domains and the JR versions. The MCIDs were greater than the corresponding MDC95 for the JR version and all HOOS domains except ADL, for which the MDC95 was 16 and the anchor MCID was 14 (Table 3).

The MCIDs calculated using the anchor-based ROC were greater than the corresponding MDC80, MDC90, and MDC95 for the JR version and all domains of the KOOS except symptom domain (Table 3). For this domain, the MCID was smaller than the MDC at every confidence level.

Substantial Clinical Benefit (SCB) Thresholds

We determined thresholds for a substantial clinical benefit for the HOOS and KOOS domains, the HOOS, JR, and the

KOOS, JR among patients who underwent THA or TKA. For THA, the substantial clinical benefit ranged from 24 to 36 and exceeded the corresponding MDCs and MCIDs for all HOOS domains and the JR version.

For TKA, the substantial clinical benefit ranged from 15 to 23 across KOOS domains and the JR version. The substantial clinical benefits exceeded the corresponding MDCs and MCIDs for the JR version and all KOOS domains except the ADL domain where the MCID was 16 and the substantial clinical benefit 15 (Table 3). The substantial clinical benefit for the KOOS ADL domain surpassed the MDC80, MDC90, and MDC95.

Proportions of Patients Achieving MCID and SCB

Across the HOOS and the HOOS, JR domains, the percentage of patients who underwent THA who attained an MCID ranged from 91% to 95% using the distribution-based method, and it ranged from 67% to 92% using the anchor-based method. Furthermore, 67% to 85% of these patients attained an anchor-based substantial clinical benefit (Table 4). More patients achieved a distribution-based MCID compared with those who achieved an anchor-based MCID. As expected, more patients attained an anchor-based MCID compared with those who got a substantial clinical benefit for the symptoms and ADL domains and for the HOOS, JR. For the pain domain, an equal number of patients obtained an anchor-based MCID and a substantial clinical benefit. For the QOL domain, the percentage of patients who attained a distribution-based

Table 4. Achievement of minimal clinically important change and substantial clinical benefit 2 years after THA and TKA

THA cohort					
HOOS domains	Minimal clinically important change			Substantial clinical benefit	
	Distribution-based Percent achieving	Anchor-based		Anchor-based	
		AUC	Percent achieving	AUC	Percent achieving
Pain	94	0.721	67	0.862	67
Symptoms	91	0.690	82	0.832	76
ADL	93	0.695	89	0.859	78
QOL	92	0.760	92	0.908	85
HOOS, JR	95	0.725	86	0.879	79

TKA cohort					
KOOS domains	Minimal clinically important change			Substantial clinical benefit	
	Distribution-based Percent achieving	Anchor-based		Anchor-based	
		AUC	Percent achieving	AUC	Percent achieving
Pain	89	0.697	81	0.835	78
Symptoms	77	0.627	78	0.762	60
ADL	86	0.707	76	0.832	77
QOL	85	0.775	81	0.911	75
KOOS, JR	91	0.670	81	0.838	68

HOOS = Hip Disability and Osteoarthritis Outcome Score; AUC = area under the curve; ADL = activities of daily living; QOL = quality of life; JR = joint replacement form of the HOOS and KOOS surveys; KOOS = Knee Injury and Osteoarthritis Outcome Score.

MCID was equal to the proportion who achieved an anchor-based MCID.

Across KOOS domains and the KOOS, JR, the percentage of patients undergoing TKA who achieved an MCID ranged from 77% to 91% using the distribution-based method; it ranged from 76% to 81% using the anchor-based method. In addition, 60% to 78% achieved an anchor-based substantial clinical benefit (Table 4). More patients achieved a distribution-based MCID than an anchor-based MCID, and more patients attained an anchor-based MCID than a substantial clinical benefit for the pain and QOL domains and the HOOS, JR. For the symptoms domain, fewer patients attained a distribution-based MCID than achieved an anchor-based MCID. For the ADL domain, fewer patients achieved an anchor-based MCID than a substantial clinical benefit.

Discussion

The potential utility of PROMs is increasingly being recognized for the assessment of care delivery and care quality, but an adequate understanding how best to measure MDC, MCID, and substantial clinical benefit for joint-specific PROMs such as HOOS and KOOS has not been adequately established. We performed a comprehensive evaluation of the properties of the MCID and substantial

clinical benefit for the HOOS, KOOS, and JR versions within a large CJR-eligible Medicare population. Consistent with previous reports, we found large variations in calculated MDCs and MCIDs depending on the methods used and assumptions made [9, 40]. Furthermore, highly variable percentages of patients achieved these thresholds depending on the methodology and the domain being evaluated. Critically, we found that MCIDs calculated using the distribution-based method are smaller than the corresponding MDCs, suggesting that these MCIDs are not valid estimates and may overestimate clinical benefit. Clinicians and policymakers should consider these findings in the process of developing an appropriate method to use HOOS, KOOS, and their JR versions to gauge patients' clinical improvement, to assess quality standards for public reporting, or to determine the appropriateness of procedures.

Limitations

There are some limitations to our study. The generalizability of our study sample is one potential limitation. Our study sample was from a single, high-volume institution, which may not be nationally representative. As an urban, tertiary care center, our institution receives referrals from near and far. Our population may be more economically

advantaged than Medicare cohorts recruited at smaller hospitals in less urban areas. The followup period we chose is another potential limitation. The MDCs, MCIDs, and substantial clinical benefits reported in this study are based on a 2-year followup, and although they are likely comparable to those calculated using 1-year followup [38], they may not be generalizable to those calculated using shorter followup [14, 15].

Like with any study requiring patient followup with PROMs, our completion rates must be sufficiently robust to minimize potential reporting bias. Although the study patients included in this analysis represent a minority of our Medicare patients, we had to make several judgments regarding who was eligible for followup. Patients undergoing surgery revision or an additional primary joint replacement before their 2-year followup were excluded, because it would be impossible to know whether the PROM scores reported at 2 years were influenced by their additional surgery. Therefore, we were limited to patients who did not have additional arthroplasty before their 2-year followup. Furthermore, the anchor-based approach is only as good as the anchor used. We have previously used our HSS Satisfaction Survey QOL question as an anchor for a study assessing factors associated with satisfaction after TKA [19]. That study population included all primary TKAs rather than only Medicare patients and the MCIDs were calculated for the WOMAC rather than the KOOS. We are currently working with Outcomes Measures in Rheumatology (OMERACT; <https://omeract.org/>) researchers to determine whether the HSS Satisfaction Survey would be a valid outcome measure in the satisfaction domain for TKA and THA.

Validity of MCID Estimates

We assessed the validity of estimates for MCID calculated using distribution- and anchor-based methods by determining whether they exceeded the MDC for the HOOS and KOOS domains, the HOOS, JR, and the KOOS, JR among patients who underwent THA or TKA. The MDC measures the precision of the selected PROM. It reflects the expected variability range if all responses were error-free. Hence, to be valid, MCIDs and substantial clinical benefits must exceed the MDC. On principle, the anchor-based ROC MCID may be considered superior to a distribution-based MCID because it relies on patient-reported improvements rather than an arbitrary sample distribution [20]. In this study, we demonstrated that the MCIDs calculated using the distribution-based approach were less than the MDC80, 90, and 95 for every domain as well as the JR versions, whereas anchor-based MCIDs exceed the MDC95 in seven of 10 domains, MDC90 in eight of 10, and MDC80 in nine of 10 (Table 3).

Although the values for the distribution-based MCIDs we reported for HOOS/KOOS are similar to those noted by other researchers [3, 4], the MDCs we described differed [26]. The KOOS MDC90s in this study (12–13 points) were substantially smaller than those from another recent study, which found MDC90s of 17 to 22 [26]. These differences can be explained by differences in sample size between the studies; our analysis was based on 2630 patients, whereas the other was based on 50 patients. Other investigators have also reported anchor-based MCIDs for the KOOS [22] and the HOOS [30] consistent with those we reported here, with one exception: the HOOS QOL domain. Paulsen et al. [30] reported an MCID of 19, whereas ours was 13. This may reflect cultural differences between study cohorts. The Paulsen cohort is based in Denmark, which was ranked the happiest country on earth in 2016 compared with the United States, which was ranked 13th [13].

Substantial Clinical Benefit Thresholds

In this study, we determined the thresholds for a substantial clinical benefit for the HOOS/KOOS domains and the HOOS/KOOS, JR among patients who underwent THA or TKA. The substantial clinical benefits for all relevant domains were uniformly large and larger for the HOOS (mean 25) than the KOOS (mean 20). This is unsurprising because THA is generally considered a more successful procedure than TKA [5, 10]. As expected, the substantial clinical benefit exceeded the MDC at all degrees of certainty (80%-95%). Substantial clinical benefit also exceeded MCID in all but two domains: HOOS symptoms and KOOS ADL. In both cases, the substantial clinical benefit was roughly equivalent to the ROC MCID. In all analyses, the MDCs, MCIDs, and substantial clinical benefits for the HOOS, JR and KOOS, JR were smaller than the equivalent measurements for the HOOS and KOOS domains. This likely reflects that activities queried on the JR surveys are universal movements, resulting in less variability resulting from health literacy or variations in patient behavior [16, 17, 33]. This represents the first assessment of the MDC, MCID, and substantial clinical benefit for the JR versions to our knowledge, so no comparison to other studies is possible.

Proportions of Patients Achieving MCID and SCB

We assessed how many patients who underwent THA and TKA achieved an MCID for the HOOS/KOOS domains and HOOS/KOOS, JR based on distribution-based and anchor-based methods. In addition, we examined how many patients achieved substantial clinical benefit with the

anchor-based method. Overall, the percentage of patients who achieved MCID and a substantial clinical benefit after THA and TKA in our cohort was high, consistent with other measures such as patient satisfaction, which reflects the effectiveness of these procedures [6, 10, 18, 27]. The percentage of patients who achieved an MCID based on the distribution-based approach was generally greater than that of the corresponding anchor-based estimates (Table 4). That all MCIDs for the distribution-based approach exceeded the corresponding MDCs (Table 3) suggests that the distribution-based approach overestimates the percentage of patients who achieve an MCID. Additionally, the proportion achieving substantial clinical benefit was lower than those achieving MCID for all domains except HOOS pain and KOOS ADL.

When using PROMs for clinical evaluation, changes in patients' scores that are lower than the MCID imply treatment failure, whereas those falling between the MCID and substantial clinical benefit indicate changes that were apparent to patients with only those exceeding the substantial clinical benefit signifying success. Consequently, it is critically important that estimates of MCIDs and substantial clinical benefits be reliable and valid. Although distribution-based MCIDs can be readily calculated from any data set with PROMs collected at different points, ease of calculation does not beget valid estimates, as we demonstrated here. Furthermore, distribution-based MCIDs may not be comparable across cohorts and hence preclude comparisons. Anchor-based MCIDs have greater face and construct validity, allowing for more accurate estimates within and across cohorts. Anchors are less readily available, and hence MCIDs based on anchors are less frequently reported, but our findings argue for routinely collecting them. Additional research is needed to confirm the results we report here. In particular, adequate anchors should be investigated to confirm the superiority of anchor-based MCIDs. Such studies will help our field to reach consensus on the standard definitions for valid and reliable MCIDs and substantial clinical benefits so that they can be applied as reliable and validated measures to guide clinical care, assess provider performance, and determine appropriateness of care.

Acknowledgments We thank Chisa Hidaka MD, Naomi Roselaar, and Caroline Boyle for their assistance in the preparation of the manuscript.

References

1. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol.* 2015;68:73–79.
2. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14:109–114.
3. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. John Charnley Award: Preoperative patient-reported outcome measures predict clinically meaningful improvement in function after THA. *Clin Orthop Relat Res.* 2016;474:321–329.
4. Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res.* 2017;475:149–157.
5. Bourne RB, Chesworth B, Davis A, Mahomed N, Charron K. Comparing patient outcomes after THA and TKA: is there a difference? *Clin Orthop Relat Res.* 2010;468:542–546.
6. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KDJ. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? *Clin Orthop Relat Res.* 2010;468:57–63.
7. Browne JP, van der Meulen JH, Lewsey JD, Lamping DL, Black N. Mathematical coupling may account for the association between baseline severity and minimally important difference values. *J Clin Epidemiol.* 2010;63:865–874.
8. Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement (CJR) model quality measures, composite quality score, and pay-for-performance methodology. 1–15. Available at: <https://doi.org/10.3928/01477447-20170302-03>. Accessed August 1, 2018.
9. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7:541–546.
10. Daigle ME, Weinstein AM, Katz JN, Losina E. The cost-effectiveness of total joint arthroplasty: a systematic review of published literature. *Best Pract Res Clin Rheumatol.* 2012;26:649–658.
11. de Groot IB, Reijman M, Terwee CB, Bierma-Zeinstra SMA, Favejee M, Roos EM, Verhaar JAN. Validation of the Dutch version of the hip disability and osteoarthritis outcome score. *Osteoarthritis Cartilage.* 2007;15:104–109.
12. Glassman SD, Copay AG, Berven SH, Polly DW, Subach BR, Carreon LY. Defining substantial clinical benefit following lumbar spine arthrodesis. *J Bone Joint Surg Am.* 2008;90:1839–1847.
13. Helliwell J, Layard R, Sachs J. World Happiness Report 2016, Update (Vol 1). *The Earth Institute*, Columbia University, New York, NY, USA. Available at: <http://eprints.lse.ac.uk/47487/>. Accessed July 17, 2018.
14. Issa K, Jauregui JJ, Given K, Harwin SF, Mont MA. A prospective, longitudinal study of patient activity levels following total knee arthroplasty stratified by demographic and comorbid factors. *J Knee Surg.* 2015;28:343–347.
15. Jauregui JJ, Issa K, Cherian JJ, Harwin SF, Given K, Mont MA. Evaluation of 5-year trends in knee society scores stratified by comorbidities: a prospective, longitudinal study. *J Knee Surg.* 2016;29:84–90.
16. Lyman S, Lee Y-Y, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res.* 2016;474:1461–1471.
17. Lyman S, Lee YY, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop Relat Res.* 2016;474:1472–1482.
18. Mancuso CA, Salvati EA, Johanson NA, Peterson MG, Charlson ME. Patients' expectations and satisfaction with total hip arthroplasty. *J Arthroplasty.* 1997;12:387–396.
19. Maratt JD, Lee Y, Lyman S, Westrich GH. Predictors of satisfaction following total knee arthroplasty. *J Arthroplasty.* 2015;30:1142–1145.

20. Mcglothlin AE, Lewis RJ. Minimal clinically important difference defining what really matters to patients. *JAMA*. 2014;312:1342–1343.
21. MN Community Measurement. New Measures Help Quantify Improvement Experienced by Patients Following Knee and Spine Surgery. 2016. Available at: <http://mncm.org/new-measures-help-quantify-improvement-patients-experienced-following-knee-and-spine-surgery/>. Accessed August 1, 2018.
22. Monticone M, Ferrante S, Salvaderi S, Motta L, Cerri C. Responsiveness and minimal important changes for the knee injury and osteoarthritis outcome score in subjects undergoing rehabilitation after total knee arthroplasty. *Am J Phys Med Rehabil*. 2013;92:864–870.
23. National Health Service. Finalised Patient Reported Outcome Measures (PROMs) in England, April 2015 to March 2016. 2017. Available at: <http://digital.nhs.uk/catalogue/PUB21189/final-proms-eng-apr14-mar15-fin-report>. Accessed July 17, 2018.
24. National Quality Forum. Patient reported outcomes (PROs) in performance measurement. 2013. Available at: https://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx. Accessed July 17, 2018.
25. National Quality Forum. NQF: Quality positioning system. 2018. Available at: <http://www.qualityforum.org/QPS/QPSTool.aspx?m=1254&e=1#qpsPageState=%7B%22TabType%22%3A1,%22TabContentType%22%3A2,%22ItemsToCompare%22%3A%5B%5D,%22StandardID%22%3A445,%22EntityTypeID%22%3A1,%22PortfolioID%22%3A3971%7D>. Accessed August 1, 2018.
26. Naylor JM, Hayen A, Davidson E, Hackett D, Harris IA, Kamalaseena G, Mittal R. Minimal detectable change for mobility and patient-reported tools in people with osteoarthritis awaiting arthroplasty. *BMC Musculoskelet Disord*. 2014;15:235.
27. Neuprez A, Delcour J-P, Fatemi F, Gillet P, Crielaard J-M, Bruyere O, Reginster J-Y. Patients' expectations impact their satisfaction following total hip or knee arthroplasty. *PLoS One*. 2016;11:e0167911.
28. Nilsson AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)—validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003;4:1–8.
29. Nwachukwu BU, Chang B, Fields K, Rebolledo BJ, Nawabi DH, Kelly BT, Ranawat AS. Defining the 'substantial clinical benefit' after arthroscopic treatment of femoroacetabular impingement. *Am J Sports Med*. 2017;45:1297–1303.
30. Paulsen A, Roos EM, Pedersen AB, Overgaard S. Minimal clinically important improvement (MCII) and patient-acceptable symptom state (PASS) in total hip arthroplasty (THA) patients 1 year postoperatively. *Acta Orthop*. 2014;85:39–48.
31. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating minimal clinically important differences in systemic lupus erythematosus. *Arthritis Res Ther*. 2015;17:143.
32. Roos EM, Lohmander LS. The knee injury and osteoarthritis outcome score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1:1–8.
33. Roos EM, Toksvig-Larsen S. Knee injury and osteoarthritis outcome score (KOOS)—validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes*. 2003;1:17.
34. Ryan A, Tompkins C. Efficiency and value in healthcare: linking cost and quality measures. *Natl Qual Forum*. 2014;1–64.
35. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. *J Rheumatol*. 2014;41:509–515.
36. Theodore BR. Methodological problems associated with the present conceptualization of the minimum clinically important difference and substantial clinical benefit. *Spine J*. 2010;10:507–509.
37. van der Roer N, Ostelo RWJG, Bekkering GE, van Tulder MW, de Vet HCW. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine (Phila Pa 1976)*. 2006;31:578–582.
38. Williams DP, Blakey CM, Hadfield SG, Murray DW, Price AJ, Field RE. Long-term trends in the Oxford knee score following total knee replacement. *Bone Joint J*. 2013;95:45–51.
39. Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther*. 2012;20:160–166.
40. Zannikos S, Lee L, Smith HE. Minimum clinically important difference and substantial clinical benefit: does one size fit all diagnoses and patients? *Semin Spine Surg*. 2014;26:8–11.